Regression in random design and wavelet block thresholding

Christophe Chesneau

Regression with Random Design and Wavelet Block Thresholding

Christophe Chesneau
Laboratoire de Probabilités et Modèles Aléatoires,
CNRS-UMR 7599, Université Paris VI, UFR de Mathématiques,
175 rue de Chevaleret F-75013 Paris, France.
E-mail: christophe.chesneau4@wanadoo.fr
Stat-page: http://www.chesneau-stat.com
15 October 2006

Abstract

In the framework of regression model with (known) random design, we prove that estimators of wavelet coefficients of the unknown regression function satisfy a strong large deviation inequality. This result can be used to show several statistical properties concerning a wavelet block thresholding estimator.

Key Words: Regression with random design, Large deviation inequality, Wavelet, Block thresholding.

AMS subject classification: Primary 62G07, Secondary 62G20.

1 Motivation

Wavelets have been shown to be a very successful tool in the framework of nonparametric function estimation. They provide adaptive estimators which enjoy good theoretical and practical properties.

In the present paper, we focus our attention on the regression model with (known) random design. We show that estimators of wavelet coefficients of the unknown regression function satisfy a strong large deviation inequality.

Thanks to this inequality, we can apply several general results established in the literature concerning the performances of a $\mathbb{L}^p$ version of an adaptive wavelet block thresholding estimator. It has been initially developed by Cai (1997, 2002) in the framework of the regression model with deterministic equispaced data. Among the numerous consequences of our inequality, we can show that the considered BlockShrink construction provides adaptive confidence intervals under the local $\mathbb{L}^p$ risk and that it achieves (near) minimax rates of convergence over Besov balls under the global $\mathbb{L}^p$ risk.

The paper is organized as follows. Section 2 describes wavelet bases. Section 3 presents the model and the main result of the paper. Some applications are described in Section 4. Section 5 contains a detailed proof of the main result.

2 Wavelets

We consider an orthonormal wavelet basis generated by dilation and translation of a compactly supported "father" wavelet ϕ and a compactly supported "mother" wavelet ψ. For the purposes of this paper, we use the periodized wavelets bases on the unit interval. Let us set

$$\phi_{j,k} = 2^{j/2}\phi(2^j x - k), \qquad \psi_{j,k} = 2^{j/2}\psi(2^j x - k).$$

And let us denote the periodized wavelets by $\phi_{j,k}^{per}(x) = \sum_{l\in\mathbb{Z}} \phi_{j,k}(x-l)$, $\psi_{j,k}^{per}(x) = \sum_{l\in\mathbb{Z}} \psi_{j,k}(x-l)$, $x \in [0,1]$. There exists an integer τ such that the collection $\zeta = \{\phi_{j,k}^{per}(x),\ k = 0, ..., 2^{\tau}-1;\ \psi_{j,k}^{per}(x),\quad j = \tau, ..., \infty,\ k = 0, ..., 2^j - 1\}$ constitutes an orthonormal basis of $\mathbb{L}^2([0,1])$. The superscript "*per*" will be suppressed from the notations for convenience.

For any integer $l \geq \tau$, a square-integrable function on $[0,1]$ can be expanded into a wavelet series

$$f(x) = \sum_{k=0}^{2^l-1} \alpha_{l,k}\phi_{l,k}(x) + \sum_{j=l}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x),$$

where $\alpha_{j,k} = \int_0^1 f(x)\phi_{j,k}(x)dx$, $\beta_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$.
For further details about wavelets, see Meyer (1990) and Cohen et al. (1993).

Since ψ is compactly supported, the following property of concentration holds : there exists a constant $C > 0$ such that, for any $m > 0$ and any $x \in [0,1]$, we have

$$\sum_{k=0}^{2^j-1} |\psi_{j,k}(x)|^m \leqslant C 2^{jm/2}. \tag{2.1}$$

3 The model and the main result

In this study, we consider the regression model with random design described as follows : suppose that we observe n pairs of random variables i.i.d $((X_1, Y_1), ..., (X_n, Y_n))$ governed by the equation:

$$Y_i = f(X_i) + z_i, \qquad i \in \{1, ..., n\}, \tag{3.1}$$

where the z_i's are Gaussian i.i.d with mean zero, variance one and are independent of the design $(X_1, ..., X_n)$. We assume that the design is known with $X_1 \in [0, 1]$. We denote by g the density of X_1 and we assume that g is bounded from above and below. The function f is unknown and bounded from above ($\|f\|_\infty$ is supposed to be known). The goal is to estimate f from observations $((X_1, Y_1), ..., (X_n, Y_n))$.

This statistical problem has been investigated by many authors via various approaches. Numerous minimax results can be found in the book of Tsybakov (2004).

If we consider the framework of wavelet analysis, the first step to estimate f consists in estimating the associated wavelet coefficients $(\beta_{j,k})_{j,k}$. Here, we consider the following unbiased estimator of $\beta_{j,k}$:

$$\hat{\beta}_{j,k} = n^{-1} \sum_{i=1}^{n} Y_i g(X_i)^{-1} \psi_{j,k}(X_i). \tag{3.2}$$

Theorem 3.1 below shows that the estimators $(\hat{\beta}_{j,k})_k$ satisfy a strong large deviation inequality. (It is important to mention that all the constants of our study are independent of f and n.)

Theorem 3.1. *Let us consider the regression model with random design described by (3.1). Let $p \in [2, \infty)$ and L be an integer such that $(\log n)^{p/2} \leqslant L < 2(\log n)^{p/2}$. Let j_1 and j_2 be integers satisfying*

$$L \leqslant 2^{j_1} < 2L, \qquad (n/\log n)^{1/2} \leqslant 2^{j_2} < 2(n/\log n)^{1/2}.$$

For any $j \in \{j_1, ..., j_2\}$, let us set $\mathcal{A}_j = \{1, ..., 2^j L^{-1}\}$ and

$$\mathcal{B}_{j,K} = \{k \in \{0, ..., 2^j - 1\} : \quad (K-1)L \leqslant k \leqslant KL - 1\}, \qquad K \in \mathcal{A}_j.$$

Then there exist two constants $\mu_1 > 0$ and $C > 0$ such that, for any $j \in \{j_1, ..., j_2\}$, $K \in \mathcal{A}_j$ and n large enough, the estimators $(\hat{\beta}_{j,k})_k$ defined by (3.2) satisfy

$$\mathbb{P}_f^n((L^{-1} \sum_{k \in \mathcal{B}_{j,K}} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geqslant \mu_1 2^{-1} n^{-1/2}) \leqslant C n^{-p}. \quad (3.3)$$

The proof of Theorem 3.1 uses several well-known large deviation inequalities as the Talagrand inequality, the Cirelson inequality and the Hoeffding inequality.

Remark 3.1. *Let us adopt the statistical framework of Theorem 3.1. For any $k \in \{0, ..., 2^j - 1\}$, there exist two constants $\mu_2 > 0$ and $C > 0$, and $K^* \in \mathcal{A}_j$ such that $k \in \mathcal{B}_{j,K^*}$ and*

$$\begin{aligned} &\mathbb{P}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}| \geqslant \mu_2 2^{-1} \sqrt{(\log n/n)}) \\ \leqslant\ \ &\mathbb{P}_f^n((L^{-1} \sum_{m \in \mathcal{B}_{j,K^*}} |\hat{\beta}_{j,m} - \beta_{j,m}|^p)^{1/p} \geqslant \mu_1 2^{-1} n^{-1/2}) \leqslant C n^{-p}. \end{aligned}$$

Therefore, the inequality (3.3) is stronger than an usual large deviation inequality generally used to investigate the (minimax) performances of the hard thresholding procedure introduced by Donoho and Johnstone (1995). See, for instance, (Kerkyacharian and Picard, 2000, Theorems 5.1 and 5.2).

Finally, let us notice that the estimator $\hat{\beta}_{j,k}$ satisfies a conventional moments condition.

Remark 3.2. *If we use the Rosenthal inequality, one can show that there exists a constant $C > 0$ such that the estimator $\hat{\beta}_{j,k}$ defined by (3.2) satisfies the following moments inequality :*

$$\mathbb{E}_f^n(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}) \leqslant C n^{-p}, \quad (3.4)$$

(see, for instance, (Kerkyacharian and Picard, 2005, Proof of Proposition 3)).

4 Some applications

In the framework of the regression model with (known) random design (3.1), the large deviation inequality proved in Theorem (3.1) can be applied to show several statistical properties concerning the following adaptive estimator of f :

$$\hat{f}_n(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k}\phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2} \sum_{K\in\mathcal{A}_j} \sum_{k\in\mathcal{B}_{j,K}} \hat{\beta}_{j,k} 1_{\left\{\hat{b}_{j,K}\geqslant \mu_1 n^{-1/2}\right\}} \psi_{j,k}(x), \quad (4.1)$$

where j_1, j_2, $\hat{\beta}_{j,k}$, $\mathcal{A}_j$, $\mathcal{B}_{j,K}$ and μ_1 are defined in Theorem 3.1, the estimator $\hat{\alpha}_{j,k}$ is defined by $\hat{\alpha}_{j,k} = n^{-1}\sum_{i=1}^{n} Y_i g(X_i)^{-1}\phi_{j,k}(X_i)$ and $\hat{b}_{j,K}$ is the normalized l_p-norm of the estimators $(\hat{\beta}_{j,k})_{k\in\mathcal{B}_{j,K}}$ i.e:

$$\hat{b}_{j,K} = (L^{-1} \sum_{k\in\mathcal{B}_{j,K}} |\hat{\beta}_{j,k}|^p)^{1/p}.$$

This construction is a $\mathbb{L}^p$ version of the BlockShrink estimator adapted to our statistical problem. Such estimator has been introduced by Cai (1997, 2002) in the framework of the regression model with deterministic equispaced data. Two consequences of the moments inequality (3.4) and the large deviation inequality (3.3) described in Theorem 3.1 are briefly described below.

- *Adaptive confidence intervals :* if we apply a result proved by Picard and Tribouley (2000, Proposition 1) then the wavelet block thresholding estimator (4.1) provides adaptive confidence intervals around $f(x_0)$ with $x_0 \in [0,1]$ under the local $\mathbb{L}^p$ risk.

- *Optimality result :* Chesneau (2006, Theorem 4.2) determines the rates of convergence achieved by the wavelet block thresholding estimator (4.1) over Besov balls under the global $\mathbb{L}^p$ risk. More precisely, by considering the Besov balls $B^s_{\pi,r}(M)$ defined by :

$$B^s_{\pi,r}(M) = \{f\in\mathbb{L}^\pi([0,1]);\ (\sum_{j=\tau-1}^{\infty} [2^{j(s+1/2-1/\pi)}(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi)^{1/\pi}]^r)^{1/r} \leqslant M\},$$

(with the usual modification if $r=\infty$) we can set the following theorem

Theorem 4.1 (Application of Theorem 4.2 proved by Chesneau (2006))**.** *Let us consider the regression model with random design (3.1). Let $p\in[2,\infty[$. Let us consider the estimator $\hat{f}_n$ defined by (4.1). Then* *there exists a constant* $C > 0$ *such that, for any* $\pi \in [1, \infty]$, $r \in [1, \infty]$, $s \in]1/\pi + 1/2, \infty)$ *and* n *large enough, we have*

$$\sup_{f \in B^s_{\pi,r}(M)} \mathbb{E}^n_f\left(\int_0^1 |\hat{f}_n(x) - f(x)|^p dx\right) \leqslant C\varphi_n,$$

where

$$\varphi_n = \begin{cases} n^{-\alpha_1 p}(\log n)^{\alpha_1 p 1_{\{p>\pi\}}}, & \text{when} \quad \epsilon > 0, \\ (\log n/n)^{\alpha_2 p}(\log n)^{(p-\pi/r)_+ 1_{\{\epsilon=0\}}}, & \text{when} \quad \epsilon \leqslant 0, \end{cases}$$

with $\alpha_1 = s/(2s+1)$, $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$ *and* $\epsilon = \pi s + 2^{-1}(\pi - p)$.

These rates of convergence are minimax and better than those achieved by the well-known hard thresholding estimator. Theorem 4.1 can be viewed as a generalization of a result proved by Chicken (2003, Theorem 2) for the uniform design, the $\mathbb{L}^2$ risk and the Hölder balls $B^s_{\infty,\infty}(M)$.

5 Proofs

First of all, let us recall the Talagrand inequality and the Cirelson inequality.

Lemma 5.1 (Talagrand (1994))**.** *Let* $(V_1, ..., V_n)$ *be i.i.d random variables and* $(\epsilon_1, ..., \epsilon_n)$ *be independent Rademacher variables, also independent of* $(V_1, ..., V_n)$*. Let* $\mathcal{F}$ *be a class of functions uniformly bounded by* T*. Let* $r_n : \mathcal{F} \to \mathbb{R}$ *be the operator defined by :*

$$r_n(h) = n^{-1}\sum_{i=1}^n h(V_i) - \mathbb{E}(h(V_1)).$$

Suppose that $\sup_{h\in\mathcal{F}} Var(h(V_1)) \leqslant v$ *and* $\mathbb{E}(\sup_{h\in\mathcal{F}} \sum_{i=1}^n \epsilon_i h(V_i)) \leqslant nH$*. Then, there exist two constants* $C_1 > 0$ *and* $C_2 > 0$ *such that, for any* $t > 0$*, we have :*

$$\mathbb{P}(\sup_{h\in\mathcal{F}} r_n(h) \geqslant t + C_2 H) \leqslant \exp(-nC_1\left(t^2v^{-1} \wedge tT^{-1}\right)).$$

Lemma 5.2 (Cirelson et al. (1976)). *Let $\mathcal{D}$ be a subset of $\mathbb{R}$. Let $(\eta_t)_{t\in\mathcal{D}}$ be a centered Gaussian process. Suppose that $\mathbb{E}(\sup_{t\in\mathcal{D}}\eta_t) \leqslant N$ and $\sup_{t\in\mathcal{D}} Var(\eta_t) \leqslant Q$. Then, for any $x>0$, we have*

$$\mathbb{P}(\sup_{t\in\mathcal{D}}\eta_t \geqslant x+N) \leqslant \exp(-x^2/(2Q)). \tag{5.1}$$

We are now in position to prove Theorem 3.1. In the following proofs, C represents a constant which may be different from one term to the other. We suppose that n is large enough.

Proof of Theorem 3.1. By the definiton of $\hat{\beta}_{j,k}$, we have the following decomposition

$$\hat{\beta}_{j,k} - \beta_{j,k} = A_{j,k} + B_{j,k}$$

where

$$A_{j,k} = n^{-1}\sum_{i=1}^{n} f(X_i)g(X_i)^{-1}\psi_{j,k}(X_1) - \mathbb{E}_f^n(f(X_1)g(X_1)^{-1}\psi_{j,k}(X_1)),$$

$$B_{j,k} = n^{-1}\sum_{i=1}^{n} g(X_i)^{-1}\psi_{j,k}(X_i)z_i.$$

By the l_p Minkowski inequality, for any $\mu>0$, we have

$$\mathbb{P}_f^n((L^{-1}\sum_{k\in\mathcal{B}_{j,K}} |\hat{\beta}_{j,k}-\beta_{j,k}|^p)^{1/p} \geqslant 2^{-1}\mu n^{-1/2}) \leqslant \mathcal{U}+\mathcal{V},$$

where

$$\mathcal{U} = \mathbb{P}_f^n((L^{-1}\sum_{k\in\mathcal{B}_{j,K}} |A_{j,k}|^p)^{1/p} \geqslant 4^{-1}\mu n^{-1/2}),$$

$$\mathcal{V} = \mathbb{P}_f^n((L^{-1}\sum_{k\in\mathcal{B}_{j,K}} |B_{j,k}|^p)^{1/p} \geqslant 4^{-1}\mu n^{-1/2}).$$

Let us investigate separately the upper bounds of $\mathcal{U}$ and $\mathcal{V}$.

• *The upper bound for $\mathcal{U}$.* Our goal is to apply the Talagrand inequality described in Lemma 5.1. Let us consider the set $\mathcal{C}_q$ defined by $\mathcal{C}_q = \{a = (a_{j,k}) \in \mathbb{Z}^*;\ \sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^q \leqslant 1\}$ and the functions class $\mathcal{F}$ defined by $\mathcal{F} =$

$\{h;\ h(x) = f(x)g(x)^{-1}\sum_{k\in\mathcal{B}_{j,K}} a_{j,k}\psi_{j,k}(x), \quad a \in \mathcal{C}_q\}$. By an argument of duality, we have

$$(\sum_{k\in\mathcal{B}_{j,K}} |A_{j,k}|^p)^{1/p} = \sup_{a\in\mathcal{C}_q} \sum_{k\in\mathcal{B}_{j,K}} a_{j,k}A_{j,k} = \sup_{h\in\mathcal{F}} r_n(h),$$

where r_n denotes the operator defined in Lemma 5.1. Let us evaluate the parameters T, H and v of the Talagrand inequality.

First of all, notice that, for $p \geqslant 2$ (and a fortiori $q = 1 + (p-1)^{-1} \leqslant 2$), an elementary inequality of l_p norm gives $\sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^2)^{1/2} \leqslant \sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^q)^{1/q} \leqslant 1$.

− *The value of* T. Let h be a function in $\mathcal{F}$. By the Hölder inequality, the assumptions of boundedness of f and g and the property of concentration (2.1), we find

$$\begin{aligned}
|h(x)| &\leqslant |f(x)||g(x)|^{-1}(\sum_{k\in\mathcal{B}_{j,K}} |\psi_{j,k}(x)|^2)^{1/2} \sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^2)^{1/2} \\
&\leqslant \|f\|_\infty \|1/g\|_\infty (\sum_{k\in\mathcal{B}_{j,K}} |\psi_{j,k}(x)|^2)^{1/2} \leqslant C2^{j/2}, \qquad x \in [0,1].
\end{aligned}$$

Hence $T = C2^{j/2}$.

− *The value of* H. The l_p-Hölder inequality and the Hölder inequality imply

$$\begin{aligned}
&\mathbb{E}^n_f(\sup_{a\in\mathcal{C}_q} \sum_{i=1}^n (\sum_{k\in\mathcal{B}_{j,K}} a_{j,k}\epsilon_i f(X_i)g(X_i)^{-1}\psi_{j,k}(X_i))) \\
&\leqslant \sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^q)^{1/q} (\sum_{k\in\mathcal{B}_{j,K}} \mathbb{E}^n_f(|\sum_{i=1}^n \epsilon_i f(X_i)g(X_i)^{-1}\psi_{j,k}(X_i)|^p))^{1/p} \\
&\leqslant (\sum_{k\in\mathcal{B}_{j,K}} \mathbb{E}^n_f(|\sum_{i=1}^n \epsilon_i f(X_i)g(X_i)^{-1}\psi_{j,k}(X_i)|^p))^{1/p}. \qquad (5.2)
\end{aligned}$$

Since $(\epsilon_1, ..., \epsilon_n)$ are independent Rademacher variables, also independent of

$\mathbb{X} = (X_1, ..., X_n)$, the Khintchine inequality yields

$$\begin{aligned}
&\mathbb{E}_f^n(|\sum_{i=1}^{n} \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i)|^p) \\
&= \mathbb{E}_f^n(\mathbb{E}_f^n(|\sum_{i=1}^{n} \epsilon_i f(X_i) g(X_i)^{-1} \psi_{j,k}(X_i)|^p | \mathbb{X})) \\
&\leqslant C\mathbb{E}_f^n(|\sum_{i=1}^{n} |f(X_i)|^2 |g(X_i)|^{-2} |\psi_{j,k}(X_i)|^2|^{p/2}) \\
&= CI. \qquad (5.3)
\end{aligned}$$

Let us consider the i.i.d random variables $(N_1, ..., N_n)$ with

$$N_i = |f(X_i)|^2 |g(X_i)|^{-2} |\psi_{j,k}(X_i)|^2, \qquad i \in \{1, ..., n\}.$$

An elementary inequality of convexity implies $I \leqslant 2^{p/2-1}(I_1 + I_2)$ where

$$I_1 = \mathbb{E}_f^n(|\sum_{i=1}^{n}(N_i - \mathbb{E}_f^n(N_1))|^{p/2}), \qquad I_2 = n^{p/2}\mathbb{E}_f^n(N_1)^{p/2}.$$

Let us analyze the upper bounds for I_1 and I_2, in turn.

– *The upper bound for* I_1*.* The Rosenthal inequality applied to $(N_1, ..., N_n)$ and the Cauchy-Schwartz inequality imply

$$\begin{aligned}
I_1 &\leqslant C(n\mathbb{E}_f^n(|N_1 - \mathbb{E}_f^n(N_1)|^{p/2}) + (n\mathbb{E}_f^n(|N_1 - \mathbb{E}_f^n(N_1)|^2))^{p/4}) \\
&\leqslant C(n\mathbb{E}_f^n(|N_1|^{p/2}) + (n\mathbb{E}_f^n(|N_1|^2))^{p/4}).
\end{aligned}$$

For any $m \geqslant 1$, $j \in \{j_1, ..., j_2\}$ and $k \in \{0, ..., 2^j - 1\}$, the assumptions of boundedness of f and g give

$$\begin{aligned}
\mathbb{E}_f^n(|N_1|^m) &= \int_0^1 |f(x)|^{2m} |g(x)|^{-2m+1} |\psi_{j,k}(x)|^{2m} dx \\
&\leqslant C2^{j(m-1)} \|\psi\|_\infty^{2m-2} \int_0^1 |\psi_{j,k}(x)|^2 dx \leqslant C2^{j_2(m-1)} \leqslant Cn^{m-1}.
\end{aligned}$$

We deduce that $I_1 \leqslant Cn^{p/2}$.

– *The upper bound for* I_2*.* Since $\mathbb{E}_f^n(N_1) \leqslant C$, we have $I_2 \leqslant Cn^{p/2}$.

Combining the obtained upper bounds for I_1 and I_2, we find

$$I \leqslant C(I_1 + I_2) \leqslant Cn^{p/2}. \tag{5.4}$$

Putting (5.2), (5.3) and (5.4) together, we see that

$$\mathbb{E}_f^n(\sup_{a\in\mathcal{C}_q}\sum_{i=1}^{n}(\sum_{k\in\mathcal{B}_{j,K}} a_{j,k}\epsilon_i f(X_i)g(X_i)^{-1}\psi_{j,k}(X_i))) \leqslant (\sum_{k\in\mathcal{B}_{j,K}} I)^{1/p} \leqslant Cn^{1/2}L^{1/p}.$$

Hence $H = Cn^{-1/2}L^{1/p}$.

$-$ *The value of v.* By the assumptions of boundedness of f and g and the orthonormality of ζ, we obtain

$$\begin{aligned}
&\sup_{h\in\mathcal{F}} Var(h(X_1))\\
&\leqslant \sup_{a\in\mathcal{C}_q}\mathbb{E}_f^n(|f(X_1)|^2|g(X_1)|^{-2}|\sum_{k\in\mathcal{B}_{j,K}} a_{j,k}\psi_{j,k}(X_1)|^2)\\
&\leqslant \|f\|_\infty^2\|1/g\|_\infty \sup_{a\in\mathcal{C}_q}\mathbb{E}_f^n(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}g(X_1)^{-1}\psi_{j,k}(X_1)\psi_{j,k'}(X_1))\\
&= C\sup_{a\in\mathcal{C}_q}\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}\int_0^1\psi_{j,k}(x)\psi_{j,k'}(x)dx\\
&= C\sup_{a\in\mathcal{C}_q}\sum_{k\in\mathcal{B}_{j,K}}|a_{j,k}|^2 \leqslant C.
\end{aligned}$$

Hence $v = C$.

Now, let us notice that, for any $j \in \{j_1, ..., j_2\}$, we have $n2^j \leqslant n2^{j_2} \leqslant 2n^{3/2}(\log n)^{-1/2}$. Since $(\log n)^{1/2} \leqslant L^{1/p} < 2^{1/p}(\log n)^{1/2}$, for $t = 8^{-1}\mu L^{1/p}n^{-1/2}$ we have

$$\left(t^2v^{-1} \wedge tT^{-1}\right) \geqslant C\left(\mu^2(\log n/n) \wedge \mu\sqrt{(\log n/(n2^j))}\right) \geqslant C\mu^2(\log n/n).$$

So, for μ large enough and $t = 8^{-1}\mu L^{1/p}n^{-1/2}$, the Talagrand inequality yields

$$\begin{aligned}
\mathcal{U} &= \mathbb{P}_f^n((L^{-1}\sum_{k\in\mathcal{B}_{j,K}}|A_{j,k}|^p)^{1/p} \geqslant 4^{-1}\mu n^{-1/2})\\
&\leqslant \mathbb{P}_f^n((L^{-1}\sum_{k\in\mathcal{B}_{j,K}}|A_{j,k}|^p)^{1/p} \geqslant 8^{-1}\mu n^{-1/2} + Cn^{-1/2})\\
&\leqslant \mathbb{P}_f^n(\sup_{h\in\mathcal{F}} r_n(h) \geqslant t + C_2H)\\
&\leqslant \exp(-nC_1\left(t^2v^{-1} \wedge tT^{-1}\right)) \leqslant \exp(-nC\mu^2(\log n/n)) \leqslant n^{-p}.
\end{aligned}$$

We obtain the desired upper bound for $\mathcal{U}$.

• *The upper bound for* $\mathcal{V}$. Our goal is to apply the Cirelson inequality described in Lemma 5.2. Let us consider the set C_q defined by $\{a = (a_{j,k}) \in \mathbb{Z}^*;\ \sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^q \leqslant 1\}$ and the process $\mathcal{Z}(a)$ defined by $\mathcal{Z}(a) = \sum_{k\in\mathcal{B}_{j,K}} a_{j,k}B_{j,k}$. Let us notice that, conditionally to $\mathbb{X} = (X_1, ..., X_n)$, $\mathcal{Z}(a)$ is a gaussian centered process. Moreover, by an argument of duality, we have

$$\sup_{a\in\mathcal{C}_q} \mathcal{Z}(a) = \sup_{a\in\mathcal{C}_q} \sum_{k\in\mathcal{B}_{j,K}} a_{j,k}B_{j,k} = (\sum_{k\in\mathcal{B}_{j,K}} |B_{j,k}|^p)^{1/p}.$$

Now, let us investigate separately the upper bounds for $\mathbb{E}^n_f(\sup_{a\in\mathcal{C}_q} \mathcal{Z}(a)|\mathbb{X})$ and $\sup_{a\in\mathcal{C}_q} Var^n_f(\mathcal{Z}(a)|\mathbb{X})$.

− *The upper bound for* $\mathbb{E}^n_f(\sup_{a\in\mathcal{C}_q} \mathcal{Z}(a)|\mathbb{X})$. Let us consider the set $\mathcal{B}_\mu$ defined by

$$\mathcal{B}_\mu = \{|n^{-1}\sum_{i=1}^n g(X_i)^{-1}|\psi_{j,k}(X_i)|^2 - 1| \geqslant \mu\}.$$

Let us work on the set $\mathcal{B}^c_\mu$, the complementary of $\mathcal{B}_\mu$. By the Jensen inequality, the fact that $\mathcal{Z}(a) \mid \mathbb{X} \sim \mathcal{N}(0, n^{-2}\sum_{i=1}^n |g(X_i)|^{-2}|\psi_{j,k}(X_i)|^2)$ and the assumptions of boundedness made on g, we find

$$\begin{aligned}
\mathbb{E}^n_f(\sup_{a\in\mathcal{C}_q} \mathcal{Z}(a)|\mathbb{X}) &\leqslant (\sum_{k\in\mathcal{B}_{j,K}} \mathbb{E}^n_f(|B_{j,k}|^p|\mathbb{X}))^{1/p}\\
&= C[\sum_{k\in\mathcal{B}_{j,K}} (n^{-2}\sum_{i=1}^n |g(X_i)|^{-2}|\psi_{j,k}(X_i)|^2)^{p/2}]^{1/p}\\
&\leqslant C\|1/g\|_\infty[\sum_{k\in\mathcal{B}_{j,K}} (n^{-2}\sum_{i=1}^n g(X_i)^{-1}|\psi_{j,k}(X_i)|^2)^{p/2}]^{1/p}\\
&\leqslant Cn^{-1/2}[\sum_{k\in\mathcal{B}_{j,K}} (n^{-1}\sum_{i=1}^n g(X_i)^{-1}|\psi_{j,k}(X_i)|^2 - 1 + 1)^{p/2}]^{1/p}\\
&\leqslant Cn^{-1/2}[\sum_{k\in\mathcal{B}_{j,K}} (\mu+1)^{p/2}]^{1/p} \leqslant C(\mu+1)^{1/2}L^{1/p}n^{-1/2}.
\end{aligned}$$

Hence $N = N(\mathbb{X}) = C(\mu+1)^{1/2}L^{1/p}n^{-1/2}$.

– *The upper bound for* $\sup_{a\in\mathcal{C}_q} Var^n_f(\mathcal{Z}(a)|\mathbb{X})$. Let us define the set $\mathcal{A}_\mu$ by

$$\mathcal{A}_\mu = \{\sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}(n^{-1}\sum_{i=1}^n g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i)) - \sum_{k\in\mathcal{B}_{j,K}}|a_{j,k}|^2) \geqslant \mu\}.$$

Let us work on the set $\mathcal{A}_\mu^c$, the complementary of $\mathcal{A}_\mu$. Using the assumptions of boundedness of g, we have

$$\begin{aligned} G &= \sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}(n^{-1}\sum_{i=1}^n |g(X_i)|^{-2}\psi_{j,k}(X_i)\psi_{j,k'}(X_i))) \\ &\leqslant C[\sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}(n^{-1}\sum_{i=1}^n g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i))... \\ &\quad -\sum_{k\in\mathcal{B}_{j,K}}|a_{j,k}|^2) + \sup_{a\in\mathcal{C}_q}\sum_{k\in\mathcal{B}_{j,K}}|a_{j,k}|^2] \leqslant C(\mu+1). \end{aligned}$$

Since $\mathbb{E}^n_f(z_i z_{i'}) = 1$ if $i = i'$ and 0 otherwise, we have

$$\begin{aligned} &\sup_{a\in\mathcal{C}_q} Var^n_f(\mathcal{Z}(a)|\mathbb{X}) = \sup_{a\in\mathcal{C}_q}\mathbb{E}^n_f(|\mathcal{Z}(a)|^2|\mathbb{X}) \\ &= \sup_{a\in\mathcal{C}_q}\mathbb{E}^n_f(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}B_{j,k}B_{j,k'}|\mathbb{X}) \\ &= \sup_{a\in\mathcal{C}_q}(n^{-2}\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}\sum_{i=1}^n\sum_{i'=1}^n |g(X_i)|^{-2}\psi_{j,k}(X_i)\psi_{j,k'}(X_{i'})\mathbb{E}^n_f(z_i z_{i'})) \\ &= n^{-1}\sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}(n^{-1}\sum_{i=1}^n |g(X_i)|^{-2}\psi_{j,k}(X_i)\psi_{j,k'}(X_i))) \\ &= n^{-1}G \leqslant Cn^{-1}(\mu+1). \end{aligned}$$

Hence $Q = Q(\mathbb{X}) = Cn^{-1}(\mu+1)$.

The obtained values of N and Q will allow us to conclude. For any $x > 0$, we have

$$\begin{aligned}
&\mathbb{P}_f^n(\sup_{a\in\mathcal{C}_q}\mathcal{Z}(a)\geqslant x+C(1+\mu)^{1/2}L^{1/p}n^{-1/2})\\
&\quad= \mathbb{E}_f^n(\mathbb{P}_f^n(\sup_{a\in\mathcal{C}_q}\mathcal{Z}(a)\geqslant x+C(1+\mu)^{1/2}L^{1/p}n^{-1/2}|\mathbb{X})(1_{\mathcal{B}_\mu}+1_{\mathcal{B}_\mu^c}))\\
&\quad\leqslant \mathbb{P}_f^n(\mathcal{B}_\mu)+\mathbb{E}_f^n(\mathbb{P}_f^n(\sup_{a\in\mathcal{C}_q}\mathcal{Z}(a)\geqslant x+N(\mathbb{X})|\mathbb{X})). \qquad (5.5)
\end{aligned}$$

The Cirelson inequality described in Lemma 5.2 implies

$$\mathbb{E}_f^n(\mathbb{P}_f^n(\sup_{a\in\mathcal{C}_q}\mathcal{Z}(a)\geqslant x+N(\mathbb{X})|\mathbb{X}))\leqslant \mathbb{E}_f^n(\exp-(x^2/(2Q(\mathbb{X})))). \qquad (5.6)$$

Moreover, by definition of $\mathcal{A}_\mu$, we have

$$\begin{aligned}
\mathbb{E}_f^n(\exp(-x^2/(2Q(\mathbb{X})))) &= \mathbb{E}_f^n(\exp(-x^2/(2Q(\mathbb{X})))(1_{\mathcal{A}_\mu}+1_{\mathcal{A}_\mu^c}))\\
&\leqslant \mathbb{P}_f^n(\mathcal{A}_\mu)+\exp(-nx^2/(2(\mu+1))). \qquad (5.7)
\end{aligned}$$

Putting the inequalities (5.5), (5.6) and (5.7) together, for $x=8^{-1}\mu L^{1/p}n^{-1/2}$ and μ large enough, we obtain

$$\begin{aligned}
\mathcal{V} &= \mathbb{P}_f^n(\sup_{a\in\mathcal{C}_q}\mathcal{Z}(a)\geqslant 4^{-1}\mu L^{1/p}n^{-1/2})\\
&\leqslant \mathbb{P}_f^n(\sup_{a\in\mathcal{C}_q}\mathcal{Z}(a)\geqslant 8^{-1}\mu L^{1/p}n^{-1/2}+C(1+\mu)^{1/2}L^{1/p}n^{-1/2})\\
&\leqslant C[\mathbb{P}_f^n(\mathcal{A}_\mu)+\mathbb{P}_f^n(\mathcal{B}_\mu)+\exp(-C\mu^2L^{2/p}/(\mu+1))]. \qquad (5.8)
\end{aligned}$$

Lemma 5.3 below provides the upper bounds for $\mathbb{P}_f^n(\mathcal{A}_\mu)$ and $\mathbb{P}_f^n(\mathcal{B}_\mu)$.

Lemma 5.3. *For μ and n large enough, there exists a constant $C>0$ such that*

$$\max(\mathbb{P}_f^n(\mathcal{A}_\mu),\mathbb{P}_f^n(\mathcal{B}_\mu))\leqslant Cn^{-p}.$$

By the inequality (5.8), the fact that $(\log n)^{p/2}\leqslant L<2(\log n)^{p/2}$ and Lemma 5.3, for μ large enough, we have

$$\mathcal{V}\leqslant Cn^{-p}.$$

Combining the obtained upper bounds for $\mathcal{U}$ and $\mathcal{V}$, we achieve the proof of Theorem 3.1. □

Proof of Lemma 5.3. Let us investigate the upper bounds for $\mathbb{P}_f^n(\mathcal{B}_\mu)$ and $\mathbb{P}_f^n(\mathcal{A}_\mu)$.

• *The upper bound for $\mathbb{P}_f^n(\mathcal{B}_\mu)$.* First of all, notice that the random variables

$$(|\psi_{j,k}(X_1)|^2 g(X_1)^{-1}, ..., |\psi_{j,k}(X_n)|^2 g(X_n)^{-1}),$$

are i.i.d. and, since g is bounded from below, we have

$$|\psi_{j,k}(X_i)|^2 g(X_i)^{-1} \leqslant \|1/g\|_\infty \|\psi\|_\infty^2 2^j, \qquad \mathbb{E}_f^n\left(|\psi_{j,k}(X_1)|^2 g(X_1)^{-1}\right) = 1.$$

So, for any $j \in \{j_1, ..., j_2\}$, the Hoeffding inequality implies the existence of a constant $C > 0$ such that

$$\mathbb{P}_f^n(\mathcal{B}_\mu) \leqslant 2\exp\left(-Cn\mu^2 2^{-2j}\right) \leqslant 2\exp\left(-Cn\mu^2 2^{-2j_2}\right) \leqslant 2n^{-C\mu^2}.$$

We obtain the desired upper bound by taking μ large enough.

• *The upper bound for $\mathbb{P}_f^n(\mathcal{A}_\mu)$.* The goal is to apply the Talagrand inequality described in Lemma 5.1. Let us consider the set $\mathcal{C}_q$ defined by $\mathcal{C}_q = \{a = (a_{j,k}) \in \mathbb{Z}^*;\ \sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^q \leqslant 1\}$ and the functions class $\mathcal{F}'$ defined by

$$\mathcal{F}' = \{h;\ h(x) = g(x)^{-1} \sum_{k\in\mathcal{B}_{j,K}} \sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}\psi_{j,k}(x)\psi_{j,k'}(x), \quad a \in \mathcal{C}_q\}.$$

We have

$$\begin{aligned} &\sup_{a\in\mathcal{C}_q}\Big(\sum_{k\in\mathcal{B}_{j,K}} \sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}\big(n^{-1}\sum_{i=1}^n g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i)\big) - \sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^2\Big) \\ = &\sup_{h\in\mathcal{F}'} r_n(h), \end{aligned}$$

where r_n denotes the operator defined in Lemma 5.1. Thus, it suffices to determine the parameter T, H and v of the Talagrand inequality.

− *The value of T.* Let h be a function of $\mathcal{F}'$. Using the Hölder inequality, the fact that g is bounded from below and the concentration property (2.1), we find

$$|h(x)| \leqslant \|1/g\|_\infty \sum_{k\in\mathcal{B}_{j,K}} |a_{j,k}|^2 \sum_{k\in\mathcal{B}_{j,K}} |\psi_{j,k}(x)|^2 \leqslant C2^j, \qquad x \in [0,1].$$

Hence $T = C2^j$.

– *The value of H.* The l_2-Hölder inequality and the Hölder inequality imply

$$\begin{aligned}
&\mathbb{E}_f^n(\sup_{a\in\mathcal{C}_q}\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}(\sum_{i=1}^n \epsilon_i g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i)))\\
&\leqslant \sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} |a_{j,k}|^2|a_{j,k'}|^2)^{1/2}...\\
&\quad [\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} \mathbb{E}_f^n(|\sum_{i=1}^n \epsilon_i(g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i))|^2)]^{1/2}\\
&\leqslant [\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} \mathbb{E}_f^n(|\sum_{i=1}^n \epsilon_i(g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i))|^2)]^{1/2}. \qquad (5.9)
\end{aligned}$$

Since $(\epsilon_1, ..., \epsilon_n)$ are independent Rademacher variables, also independent of $(X_1, ..., X_n) = \mathbb{X}$, the Khintchine inequality and the fact that g is bounded from below imply

$$\begin{aligned}
&\mathbb{E}_f^n(|\sum_{i=1}^n \epsilon_i(g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i))|^2)\\
&= \mathbb{E}_f^n(\mathbb{E}_f^n(|\sum_{i=1}^n \epsilon_i(g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i))|^2|\mathbb{X}))\\
&\leqslant C\mathbb{E}_f^n(\sum_{i=1}^n |g(X_i)|^{-2}|\psi_{j,k}(X_i)|^2|\psi_{j,k'}(X_i)|^2)\\
&\leqslant C\|1/g\|_\infty^2 n\mathbb{E}_f^n(|\psi_{j,k}(X_1)|^2|\psi_{j,k'}(X_1)|^2). \qquad (5.10)
\end{aligned}$$

Using the property of concentration (2.1) and the inequalities (5.9) and (5.10), we find

$$\begin{aligned}
&\mathbb{E}_f^n(\sup_{a\in\mathcal{C}_q}\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}(\sum_{i=1}^n \epsilon_i g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i)))\\
&\leqslant C[n\mathbb{E}_f^n((\sum_{k\in\mathcal{B}_{j,K}} |\psi_{j,k}(X_i)|^2)^2)]^{1/2} \leqslant Cn^{1/2}2^j.
\end{aligned}$$

Hence $H = C2^j n^{-1/2}$.

– *The value of* v. Using the fact that g is bounded from below, the Hölder inequality and the property of concentration (2.1), we have

$$\begin{aligned}\sup_{h\in\mathcal{F}} Var(h(X_1)) &\leqslant \sup_{a\in\mathcal{C}_q} \mathbb{E}_f^n(|g(X_1)|^{-2}|\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}\psi_{j,k}(X_1)\psi_{j,k'}(X_1)|^2) \\ &\leqslant C\|1/g\|_\infty^2 \sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}}|a_{j,k}|^2)^2\mathbb{E}_f^n((\sum_{k\in\mathcal{B}_{j,K}}|\psi_{j,k}(X_1)|^2)^2) \\ &\leqslant C2^{2j}.\end{aligned}$$

Hence $v = C2^{2j}$.

Now, let us notice that if $t = 2^{-1}\mu$ then

$$\left(t^2v^{-1}\wedge tT^{-1}\right) \geqslant C\left(\mu^2 2^{-2j}\wedge \mu 2^{-j}\right) = C\mu^2 2^{-2j}.$$

For any $j\in\{j_1,...,j_2\}$, μ large enough and $t = 2^{-1}\mu$, the Talagrand inequality gives

$$\begin{aligned}&\mathbb{P}_f^n(\mathcal{A}_\mu) \\ &\leqslant \mathbb{P}_f^n(\sup_{a\in\mathcal{C}_q}(\sum_{k\in\mathcal{B}_{j,K}}\sum_{k'\in\mathcal{B}_{j,K}} a_{j,k}a_{j,k'}(n^{-1}\sum_{i=1}^n g(X_i)^{-1}\psi_{j,k}(X_i)\psi_{j,k'}(X_i)) - ... \\ &\qquad \sum_{k\in\mathcal{B}_{j,K}}|a_{j,k}|^2) \geqslant 2^{-1}\mu + C2^j n^{-1/2}) \leqslant \mathbb{P}(\sup_{h\in\mathcal{F}} r_n(h)\geqslant t + C_2H) \\ &\leqslant \exp(-nC_1\left(t^2v^{-1}\wedge tT^{-1}\right)) \leqslant \exp(-nC\mu^2 2^{-2j}) \\ &\leqslant \exp(-nC\mu^2 2^{-2j_2}) \leqslant n^{-p}.\end{aligned}$$

This ends the proof of Lemma 5.3. □

References

Cai, T. (1997). On adaptivity of blockshrink wavelet estimator over besov spaces. *Technical Report, 97-05, Department of Statistics, Purdue University*.

Cai, T. (2002). On block thresholding in wavelet regression: adaptivity, block size and threshold level. *Statist. Sinica*, 12(4):1241–1273.

CHESNEAU, C. (2006). Wavelet estimation via block thresholding : A minimax study under the $\mathbb{L}^p$ risk. *Available on HAL, http://hal.ccsd.cnrs.fr/*.

CHICKEN, E. (2003). Nonparametric regression on random processes and design. *Florida State University Department of Statistics, Technical Report*.

CIRELSON, B., IBRAGIMOV, I., and SUDAKOV, V. (1976). *Norm of Gaussian sample functions*. Berlin, springer verlag ed.

COHEN, A., DAUBECHIES, I., JAWERTH, B., and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24(1):54–81.

DONOHO, D. and JOHNSTONE, I. (1995). Adaptating to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

KERKYACHARIAN, G. and PICARD, D. (2000). Thresholding algorithms, maxisets and well concentrated bases. *Test*, 9(2):283–345.

KERKYACHARIAN, G. and PICARD, D. (2005). Regression in random design and warped wavelets. *Bernoulli*, 10(6):1053–1105.

MEYER, Y. (1990). *Ondelettes et Opérateurs*. Hermann, Paris.

PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann.Statist.*, 28(1):298–335.

TALAGRAND, M. (1994). Sharper bounds for gaussian and empirical processes. *Ann. Probab.*, 22:28–76.

TSYBAKOV, A. (2004). *Introduction à l'estimation nonparametrique*. Springer Verlag.